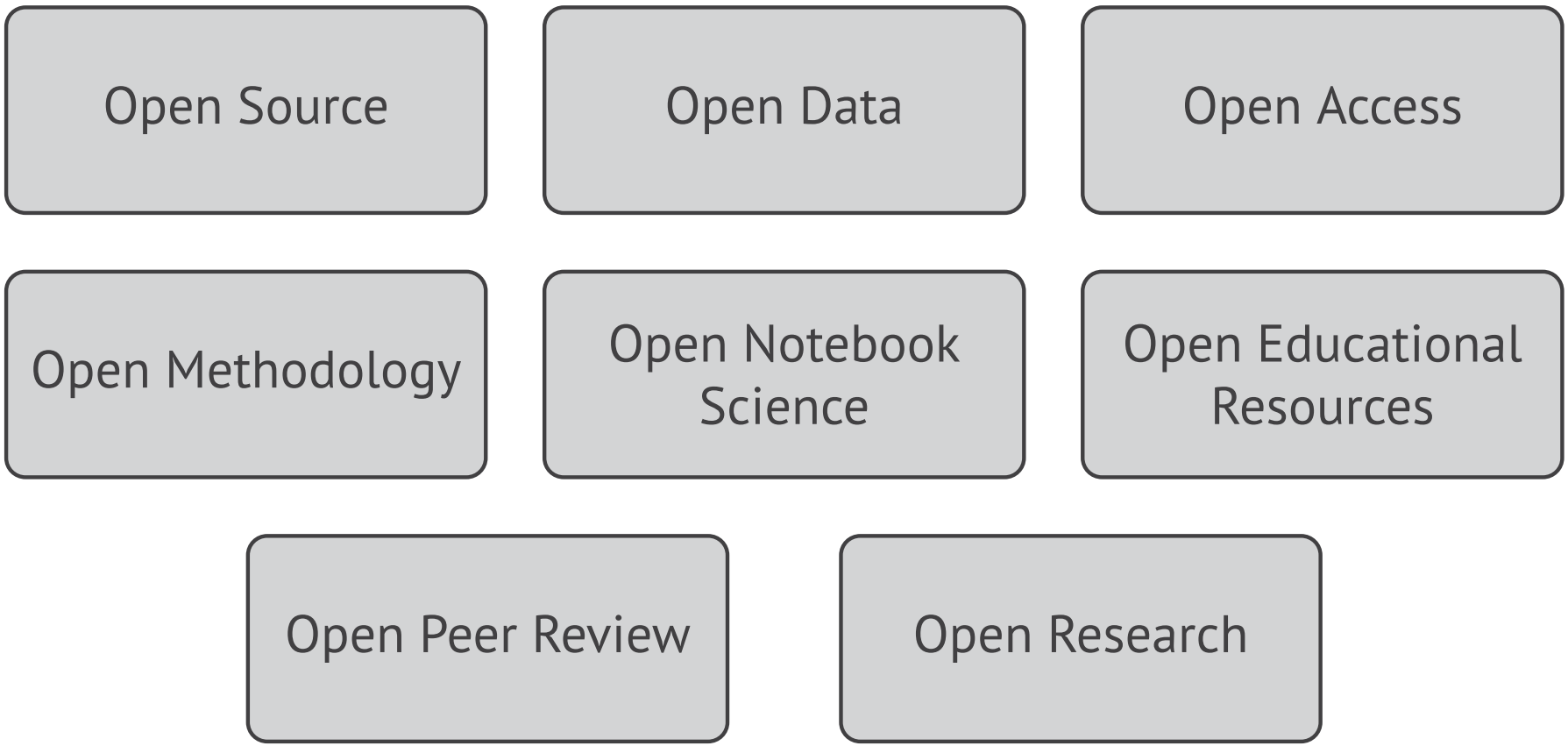
Open Science in the Two!Ears Project Experiences and Best Practices

Hagen Wierstorf [1] Fiete Winter[2] **Sascha Spors**[2]

[1] Centre for Vision, Speech and Signal Processing, University of Surrey
[2] Institute of Communications Engineering, University of Rostock

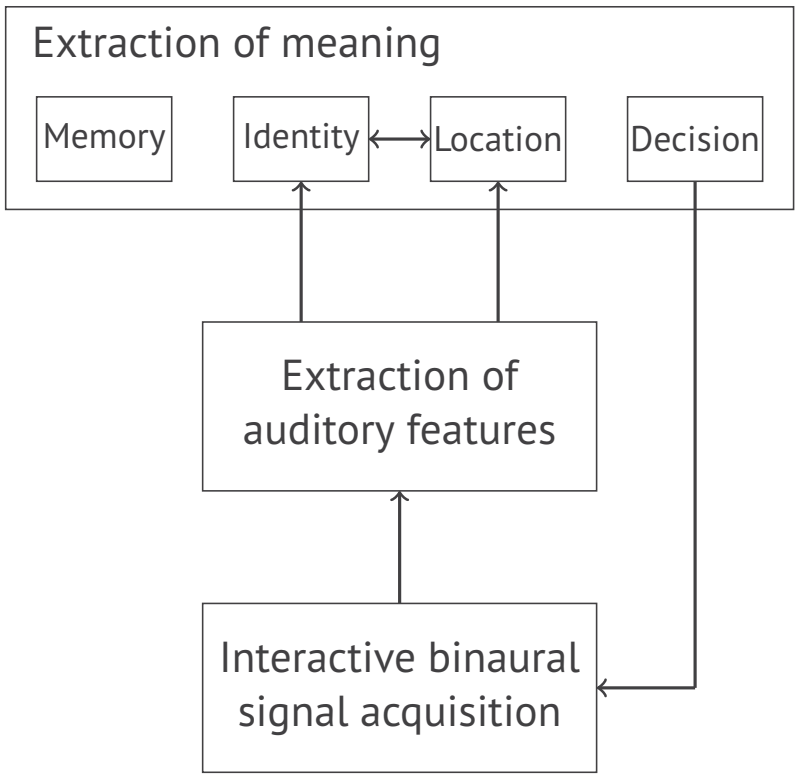
Acoustics ’17 Boston

The Elements of Open Science



- Requires systematic management of research data
- Open science by itself does not ensure the ease of reproducibility

Computational framework for modelling active exploratory listening that assigns meaning to auditory scenes



- 9 international partners
- Scenario-based development and evaluation
- Modular architecture
- Simulated/real-world input

⇒ **open source software and data**

Comparison with AMToolbox

Auditory Modeling Toolbox

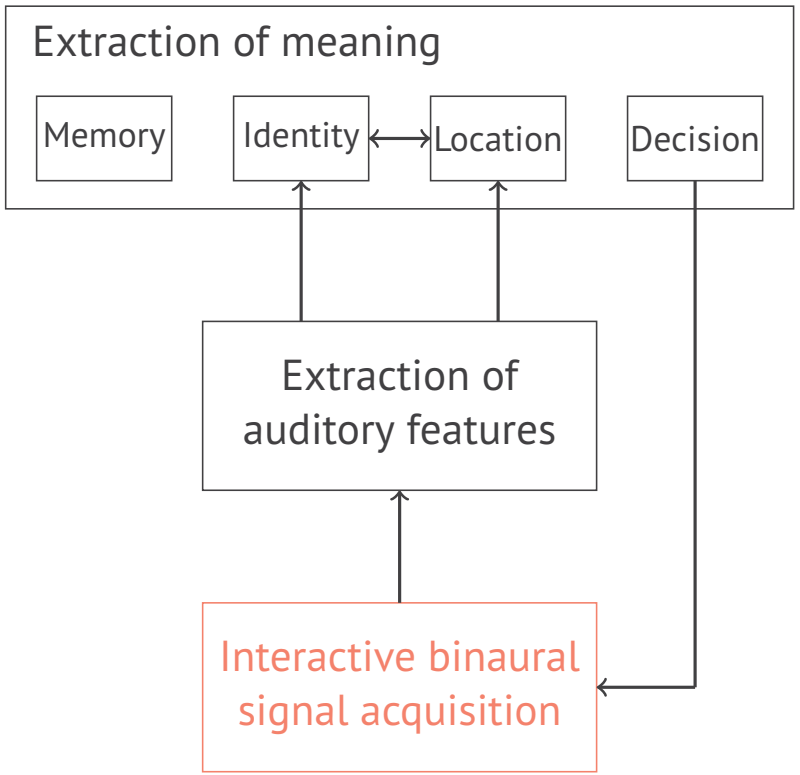
- AABBA project started in 2009: **apply binaural models**
- Source code of models rarely available
- Initiated open collection of models:
  `http://amtoolbox.sourceforge.net`

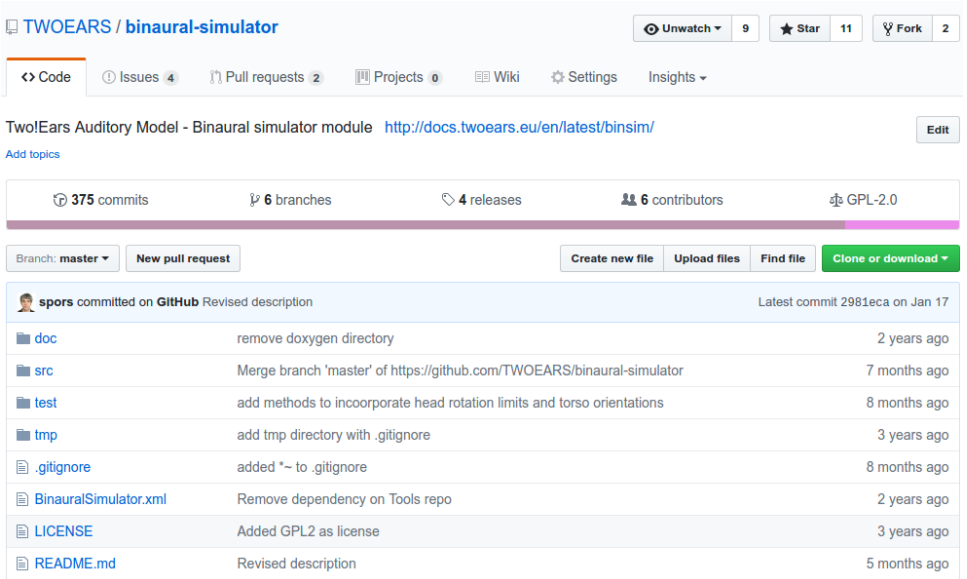
Additional features required by Two!Ears

- Block-based processing
- Clear separation in feature extraction and machine learning
- Seamless combination of different approaches
- Initiated dedicated modeling approach:
  `http://docs.twoears.eu`

Two!Ears Auditory Model

Binaural Simulator

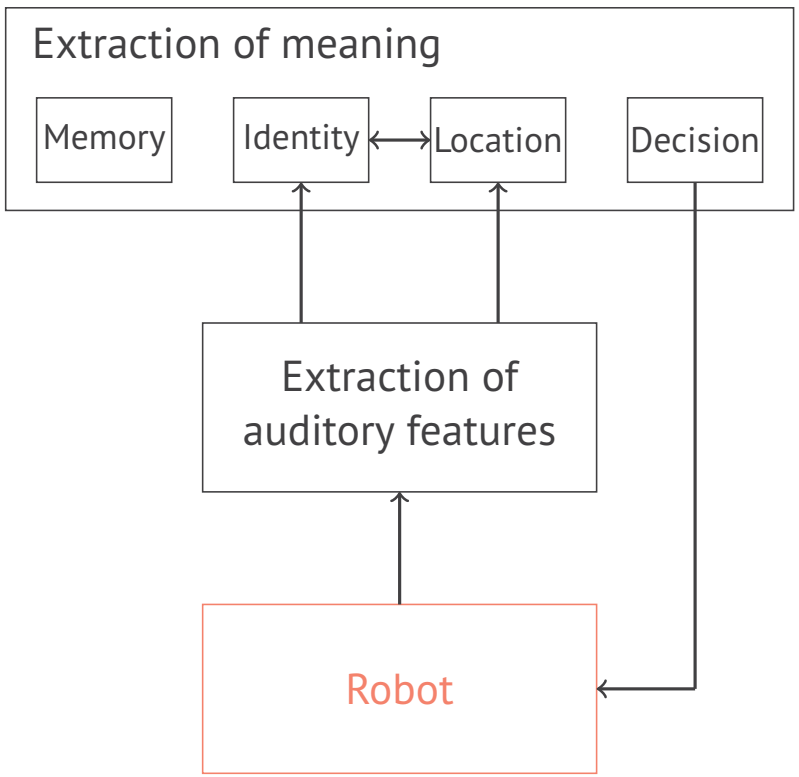


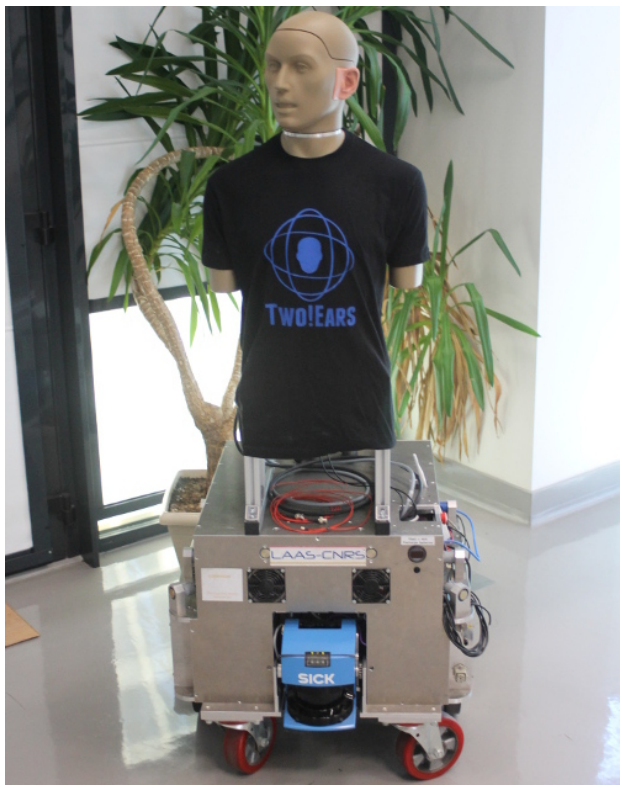


- Defined interface to other modules
- May be used standalone

Two!Ears Auditory Model

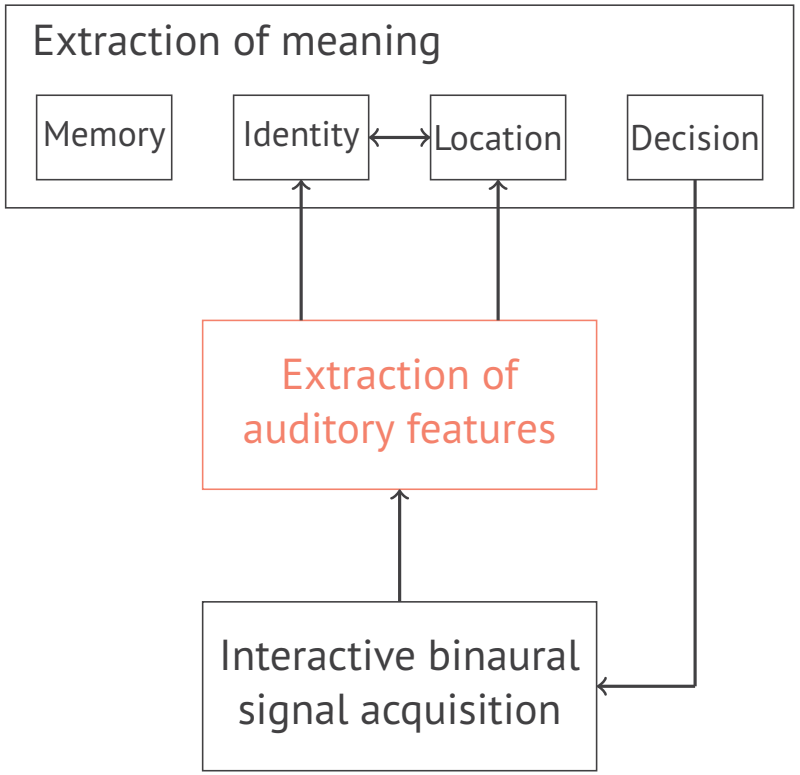
Real-World input by robot

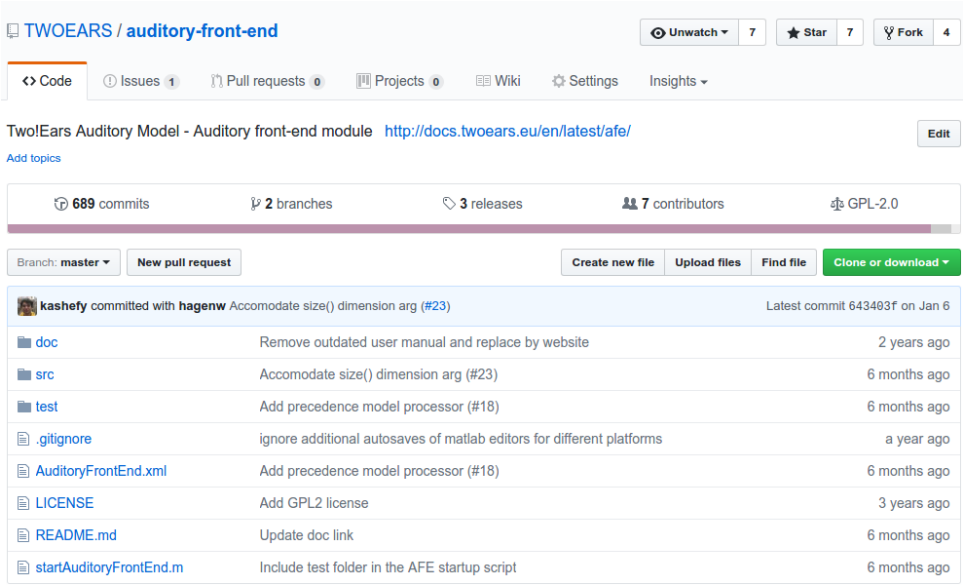




Two!Ears Auditory Model

Auditory front-end

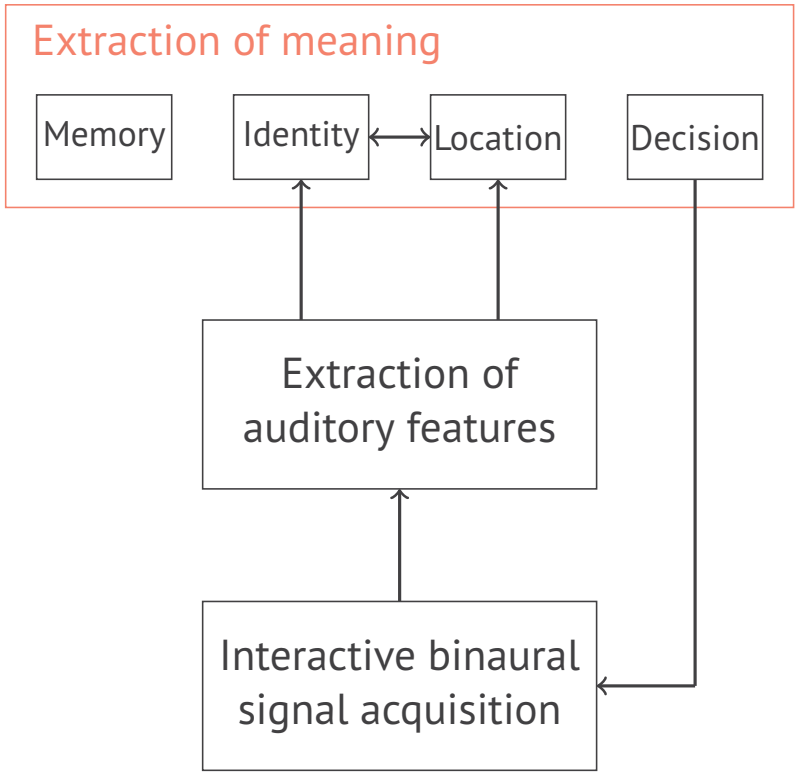




- Extracts auditory features
- May be used standalone

Two!Ears Auditory Model

Blackboard system



TWOEARS / **blackboard-system**

Unwatch 7 | Star 2 | Fork 1

Code | Issues 7 | Pull requests 0 | Projects 0 | Wiki | Settings | Insights

Two!Ears Auditory Model - Blackboard system module http://docs.twoears.eu/en/latest/blac… Edit

Add topics

1,023 commits | **9** branches | **1** release | **7** contributors | GPL-2.0

Branch: **master** | New pull request | Create new file | Upload files | Find file | Clone or download

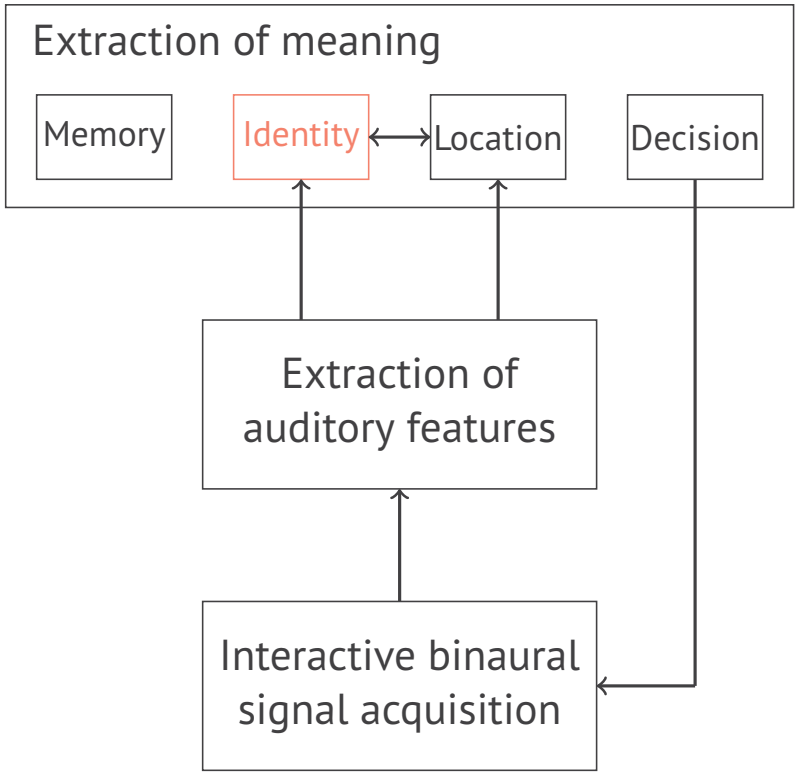
Ning Ma fixed anglular distance bug — Latest commit 17056a9 on Jan 19

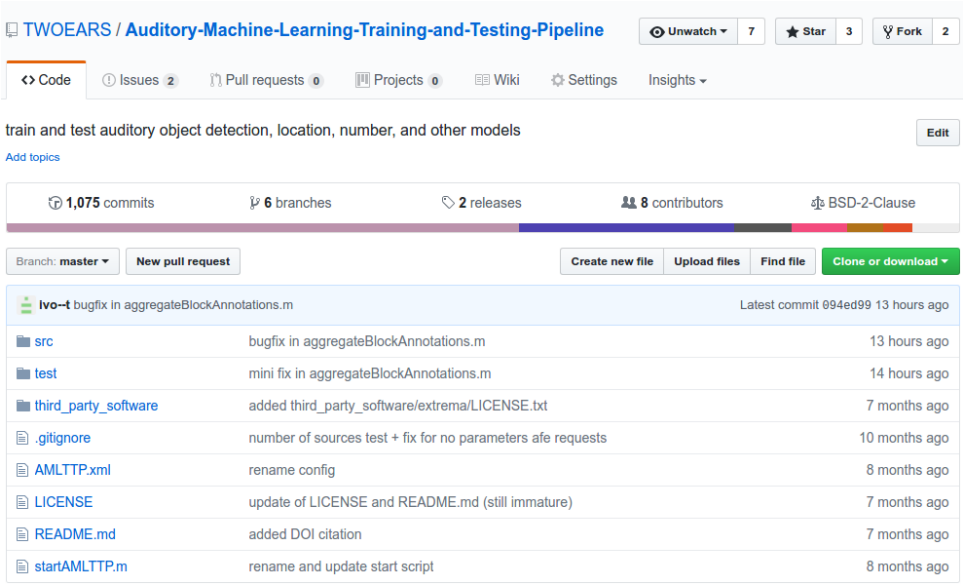
doc	created evalutaion folder for evaluation scripts	3 years ago
src	fixed anglular distance bug	5 months ago
test	Remove further GMTK stuff	a year ago
.gitignore	added GMTK binaries for MacOSX to third_party_software/gmtk-1.0.1.bin…	3 years ago
BlackboardSystem.xml	Remove Tools dependency	2 years ago
LICENSE	Add GOL2 license	2 years ago
README.md	Update doc links in README	6 months ago
startBlackboardSystem.m	typo in startBlackboardSystem.m	a year ago

- Knowledge sources for various tasks
- Scheduler

Two!Ears Auditory Model

Training and testing pipeline





- Supports training
- Trained models → database

Conclusions

Software Development

Things that worked great:

- git and github excellent tools for collaborative software development
- readthedocs good place for creating documentation

Proposal for future projects:

- Make a workshop on the topic at the beginning
- Include a software engineer for complex software projects
- Try to avoid usage of closed-source software (e.g. Matlab)
- Training on reproducible research and publications

Requirements on Data Management

- Data is collected and modified during the project:
  - HRIR/BRIR measurements for acoustic simulations
  - Training of machine learning stages
  - Listening test results
- All partners need seamless access
- Not all data can be made publicly available
- Potentially different versions of the same data set

$\Rightarrow$ Ideal solution: version control for data + rights management

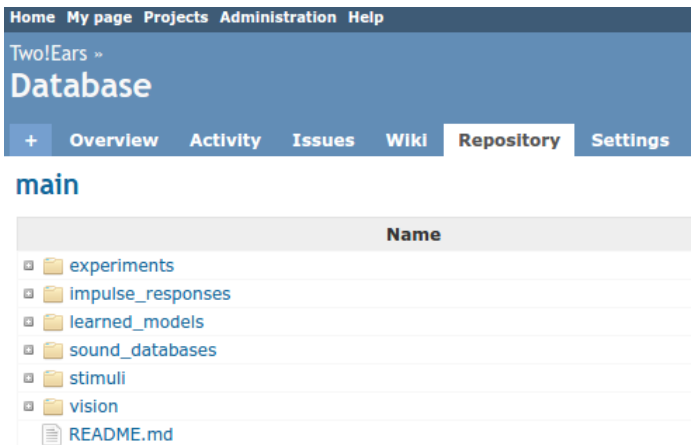
Possible Approaches to Data Management

- svn works, but branching becomes buggy
- git may produce out of memory errors on the server
- Git Large File Storage was released during the project, but lacked a working server implementation
- Similar implementations from the community, like git-media, git-annex
- Commercial providers like BitKeeper

Our Solution to Data Management

- Modified version of git-media for internal repository
- svn and Redmine for public repository
- In both cases you can download single files and subdirectories

Web-frontend



Matlab interface

```
fname = db.getFile('path/to/file');
sig = audioread(fname);
```

Conclusions

Data management

Things that worked ok:

- svn + Redmine for public data
- zenodo for releasing single data sets

Proposal for future projects:

- Avoid complicated setups (like our git-media)
- Hope for better tools

Summary

Implementation of Open Science in Two!Ears

- Internal management of research data by Redmine
- Open source software & database, extensive documentation
- Includes full version history

Some lessons learned...

- Open science requires training and qualification
- Seamless integration and usability of tools essential
- Counteracting evaluation measures

`github.com/twoears`

`github.com/spatialaudio`